Predicting the Ability to Maintain Alertness During Sleep Deprivation: The Accuracy of Subjective Evaluations

By

John A. Caldwell, Jr.

and

Peggy S. Ruyak

Aircrew Health and Performance Division

November 1997

19971218 032

DTIC QUALITY INSPECTED 4

**U.S. Army Aeromedical Research Laboratory
Fort Rucker, Alabama 36362-0577**

Reviewed:

MORRIS R. LATTIMORE, JR.
LTC, MS
Director, Aircrew Health &
   Performance Division

Released for publication:

JOHN A. CALDWELL, Ph.D.
Chairman, Scientific
   Review Committee

CHERRY L. GAFFNEY
Colonel, MC, SFS
Commanding

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS

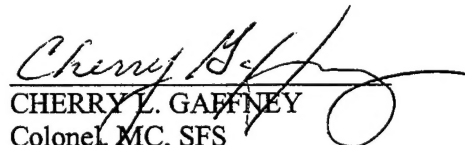| 2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release, distribution unlimited
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	

4. PERFORMING ORGANIZATION REPORT NUMBER(S) USAARL Report No. 98-05	5. MONITORING ORGANIZATION REPORT NUMBER(S)

6a. NAME OF PERFORMING ORGANIZATION U.S. Army Aeromedical Research Laboratory	6b. OFFICE SYMBOL (If applicable) MCMR-UAC	7a. NAME OF MONITORING ORGANIZATION U.S. Army Medical Research and Materiel Command

6c. ADDRESS (City, State, and ZIP Code) P.O. Box 620577 Fort Rucker, AL 36362-0577	7b. ADDRESS (City, State, and ZIP Code) Fort Detrick Frederick, MD 21702-5012

8a. NAME OF FUNDING / SPONSORING ORGANIZATION	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER

8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS

	PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
	0602787A	3M162787A879	OD	165

11. TITLE (Include Security Classification)
(U) Predicting the ability to maintain alertness during sleep deprivation: The accuracy of subjective evaluations

12. PERSONAL AUTHOR(S)
John A. Caldwell, Jr. and Peggy S. Ruyak

13a. TYPE OF REPORT Final	13b. TIME COVERED FROM TO	14. DATE OF REPORT (Year, Month, Day) 1997 November	15. PAGE COUNT 15

16. SUPPLEMENTAL NOTATION

17.	COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) Sleepiness, subjective, objective, RTSW, VAS, and POMS.
FIELD	GROUP	SUB-GROUP	
06	10		
05	08		

19. ABSTRACT (Continue on reverse if necessary and identify by block number)
To determine the extent of agreement between subjective measures of mood/sleepiness and an objective electroencephalographic (EEG)-based measure of sleepiness/alertness, 18 subjects were administered the Profile of Mood States (POMS), a Visual Analog Scale (VAS), and the Repeated Test of Sustained Wakefulness (RTSW) throughout 38-hour periods of continuous wakefulness. Multivariate analysis of variance (MANOVA) indicated that all of the measures, with the exception of VAS anxiety, were sensitive to the effects of sleep loss. Multiple regression analysis showed that 34 percent of the overall variance in RTSW scores could be accounted for by the linear combination of the subjective self-report scores, although when the data were grouped according to whether subjects were totally sleep deprived or whether they were permitted a 2-hour nap, the percentages ranged from 53 to 24 percent. A stepwise regression indicated that VAS sleepiness was the best predictor of RTSW and POMS fatigue was the next best, followed by VAS alertness. Bivariate correlations showed that the 3 best predictors (sleepiness, fatigue, and alertness) correlated -0.49, -0.47, and 0.36 with RTSW scores overall; however, bivariate correlations conducted within each cell of the design (with 18 observations per cell) generally were not (Cont.)

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Chief, Science Support Center	22b. TELEPHONE (Include Area Code) (334) 255-6907	22c. OFFICE SYMBOL MCMR-UAX-SI

DD Form 1473, JUN 86 Previous editions are obsolete. SECURITY CLASSIFICATION OF THIS PAGE

19. Abstract, Continued

significant due to low statistical power and the relatively weak relationship between subjective measures and RTSW sleep latency. A stepwise discriminant analysis conducted on the variables indicated that 58 percent of the time subjects were correctly classified into groups (based on RTSW scores) of low, medium, and high alertness. The results of this study indicate that an objective measure of sleepiness/alertness which emphasizes the maintenance of wakefulness (the RTSW) is not more strongly related to self-reports of sleepiness than tests which emphasize the ability to fall asleep, such as the Multiple Sleep Latency Test (MSLT).

Table of contents

Introduction

Various subjective questionnaires, performance tests, and electroencephalographic-(EEG) based measures have been utilized to determine the level of sleepiness in normal persons and patients with sleep disorders (Mitler and Miller, 1996). Self-reports such as the Stanford Sleepiness Scale (SSS), the Epworth Sleepiness Scale (ESS), and the Visual Analog Scale (VAS) have been used to quantify the subjective desire for sleep (Hoddes et al., 1973; Johns, 1991; and Penetar et al., 1993). These instruments are popular because of low cost, ease of administration, and repeatability, and they are utilized in a variety of settings. Performance-based measures such as the Walter Reed Performance Assessment Battery (Thorne et al., 1985), the Synthetic Work Battery (Elsmore et al., 1995), and the Wilkinson 5-Choice Test (Wilkinson, 1965) also purport to tap the effects of sleepiness, although they are used primarily in laboratory experiments. Serial reaction time and vigilance tasks appear to be most degraded by sleep loss, whereas rote learning exercises and game-like computer tests are less affected (Wilkinson, 1964). EEG-based measures are the most objective techniques for determining sleepiness and are employed in both clinical and experimental settings. These include the Alpha Attenuation Test (AAT), the Multiple Sleep Latency Test (MSLT), and variants of the MSLT including the Repeated Test of Sustained Wakefulness (RTSW) and the Maintenance of Wakefulness Test (MWT) (Alloway et al., 1997; Carskadon and Dement, 1982; Hartse, Roth, and Zorick, 1982; and Sangal et al., 1992). Of these, the MSLT is the most widely used. The MSLT is a standardized and normed procedure that quantifies the amount of sleepiness as mild, moderate, or severe (ASDA, 1992). The validity of the MSLT in differentiating levels of sleepiness has been shown in patients with sleep apnea and narcolepsy, as well as in normals undergoing sleep deprivation (Carskadon and Roth, 1994).

Each of the above strategies provides relevant information; however, they are not equivalent. For instance, although tests such as the Wilkinson Addition and Vigilance and the modified Williams Word Memory have been shown to correlate with SSS ratings (Hoddes et al., 1973), performance tests in general are subject to biases associated with task length and complexity, subject motivation, experimenter feedback, and other factors (Wilkinson, 1961; Wilkinson, 1964; Elsmore et al., 1995; Wilkinson, 1969). Also, while self-report inventories can quickly and inexpensively differentiate levels of sleepiness associated with sleep disorders and sleep deprivation, and while strong relationships between the results of the SSS and VAS have been found (Johnson et al., 1991), the results of self-reports correlate poorly with more objective techniques (i.e., MSLTs) (Johnson et al., 1991; Chervin et al., 1995; Seidel et al., 1987).

Perhaps part of the reason for low correlations is that subjective and objective strategies are measuring different dimensions of sleepiness. Introspective or behavioral techniques appear to reflect manifest sleepiness/alertness susceptible to moment-by-moment fluctuations in factors affecting arousal (such as environmental influences, activity level, physical discomfort, etc.), whereas the MSLT reflects physiological sleepiness, and thus is thought to be a more direct

1

measure of true underlying sleep tendency (Carskadon and Dement, 1982). Broughton (1992) suggests further that the MSLT may confound sleepiness with the learned ability to fall asleep. Still others believe the MSLT generally lacks sensitivity because of a "floor effect" which occurs when sleep latencies approach zero, as is the case with extremely sleepy individuals (Sugerman and Walsh, 1989). This problem may have contributed to past findings that pre- versus post-treatment MSLTs were unaffected despite improvements in nocturnal sleep quality, nocturnal respiration, and subjective estimates of sleepiness; and as result, it has been questioned whether treatments for sleepiness might be better assessed by a test which measures the maintenance of wakefulness rather than the initiation of sleep (Roth et al., 1980). If it is true that a variant of the MSLT (such as the MWT) could be more appropriate than the MSLT in some situations, perhaps it is also the case that variants of the MSLT may correlate differently with other measures of sleepiness/alertness.

Hartse, Roth and Zorick (1982) proposed a modification to the MSLT based on the premise that the ability of people to stay awake is of greatest interest when studying excessive daytime sleepiness. To tap the ability to remain awake, the RTSW was developed. This test, a variant of the MWT (Mitler et al., 1982), consists of a standard protocol in which subjects are placed in a setting identical to the one used for MSLTs, but instead of being instructed to go to sleep, they are instructed to remain awake. Hartse, Roth, and Zorick (1982) compared the RTSW and MSLT and reported that because of the instructions to remain awake, mean sleep latencies were longer in the RTSW, but regardless of instructional set, normal subjects were not able to remain awake under conditions of sleep deprivation. Subsequent research has shown that both MSLTs and RTSWs are sensitive to the effects of sleep loss (Walsh et al., 1990) and circadian variations in sleepiness/alertness (Clodore et al., 1990). These findings suggest that other tests of sleepiness may correlate about as well (or as poorly) with the RTSW as with the MSLT; however, the relationship between the RTSW and subjective ratings has not been well studied.

The present experiment evaluated the strength of the relationship between latency to sleep using the RTSW procedure and subjective ratings on two self-report inventories. Normal volunteers were tested under nonsleep-deprived and sleep-deprived conditions.

Methods

Subjects

Eighteen male subjects between the ages of 22 and 31 (mean=24.4) were tested at the United States Army Aeromedical Research Laboratory (USAARL) after signing informed consent agreements. Subjects remained awake for 36-38 continuous hours under each of three interventions--a 2-hour nap (from 2100 to 2300) initiated with 10 mg zolpidem (Znap), a 2-hour placebo nap (Pnap), and a 2-hour forced-rest period (Nonap). Data were collected during pre-deprivation and deprivation sessions.

<center>Apparatus</center>

<u>Subjective evaluations</u>

Profile of Mood States (POMS)

The POMS (McNair, Lorr, and Droppleman, 1981) was used to assess subjective reports of mood at various times throughout the day. This paper-and-pencil questionnaire consisted of 65 items which measured affect on 6 scales: tension-anxiety, depression-dejection, anger-hostility, vigor-activity, fatigue-inertia, and confusion-bewilderment. The answers were scored by hand with scoring templates.

Visual Analog Scale (VAS)

The VAS was used to measure subjective sleepiness. The VAS consisted of eight 100 mm lines centered over the adjectives "alert/able to concentrate", "anxious", "energetic", "feel confident", "irritable", "jittery/nervous", "sleepy", and "talkative" (Penetar et al., 1993). At the extremes of each line, "not at all" and "extremely" were printed respectively. Scores consisted of the distance of the subject's mark from the left end of the line (in mm).

<u>Objective evaluation</u>

Objective sleepiness was measured using the RTSW (Hartse, Roth, and Zorick, 1982) in which subjects' EEG was recorded for up to 20 minutes using a Nihon Kohden electroencephalogram (Model No. EEG-4321P) during the test to determine whether or not subjects successfully remained awake (subjects were awakened and removed from the room immediately if they fell asleep). Records were scored in terms of the number of minutes from lights out until sleep onset (up to 20 minutes).

<center>Procedure</center>

<u>Schedule</u>

Subjects were housed in the laboratory for 9 days. After a Sunday adaptation sleep night (10 hours), training began at 0900 (on Monday) and lasted until 2010 (bedtime was at 2200). On Monday, Wednesday, Friday, and Sunday nights, subjects slept for 10 hours. On Tuesday, Thursday, and Saturday (the predeprivation/intervention days), testing was conducted and subjects received one of the interventions instead of a full night's sleep. Either Pnap, Znap, or Nonap was conducted on each night (from 2100-2300). All subjects received every intervention in counterbalanced fashion. On Wednesday, Friday, and Sunday (the test days following interventions), subjects were tested from 0100 until 2010.

<center>3</center>

<u>Subjective evaluations</u>

The POMS was administered every 2 hours from 0900-1900 on predeprivation days and from 0100-1900 on deprivation days. The test took approximately 5 minutes. The VAS was administered every 2 hours from 1000-2000 on predeprivation days and from 0200-2000 on deprivation days. The VAS took approximately 5 minutes.

<u>Objective evaluation</u>

RTSWs were conducted every 2 hours from 1010-2010 on predeprivation days and from 0210-2010 on deprivation days. Each RTSW lasted up to 20 minutes, but subjects were awakened and removed from the room immediately when they fell asleep (as indicated by the presence of the first sleep spindle or K-complex). Records were scored in terms of the number of minutes from lights out until sleep onset.

Data Analysis

Each of the 18 subjects contributed data from 16 sessions (6 predeprivation and 10 deprivation) for each of the 3 interventions (Znap, Pnap, and Nonap). For overall significance tests, these data were analyzed in a repeated measures multiple analysis of variance (MANOVA). For tests of association (between RTSW and the other scores) these data were treated as independent observations (or cases), similar to the approach used by Kryger et al. (1991). This yielded 864 cases (18 subjects x 3 conditions x 16 sessions) which were analyzed in several steps. Since this approach violates the assumption of independence among observations, the net result is to decrease the error term which tends to make significance tests less stringent. To compensate for this problem (since no regression procedures exist for repeated measures designs), the alpha level for rejecting the null hypothesis was changed from 0.05 to 0.01 for all tests in which the replicates for subjects were treated as independent observations.

MANOVA

To determine whether there were overall differences among RTSW, VAS, and POMS scores as a function of treatment intervention (Znap, Pnap, and Nonap) and sleep deprivation (predeprivation versus deprivation sessions), BMDP4V was used to perform a MANOVA. Wilk's lambda likelihood ratio was the statistic chosen to evaluate whether or not a significant difference existed among the interventions. Wilk's lambda is the proportion of total variance in the vectors (which represent the multiple dependent variables) that is not explained by differences among groups. Thus, a small lambda is indicative of significant group (or intervention) differences. The nature of these differences (determined by univariate analysis of variance, simple effects, and posthoc contrasts) is not presented in the present paper; however, it is detailed in Caldwell and Caldwell (in press).

Multiple regression

To estimate a least squares linear regression between latency to stage 2 in the RTSW (the dependent variable) and the 14 POMS and VAS scores (the independent variables), BMDP1R was used. First, an overall multiple R on the data from all three conditions was computed, and then a multiple R within each separate condition was computed. In addition, a test of slopes was calculated to determine whether the relationship between RTSW and subjective indices differed as a function of condition.

Stepwise multiple regression

To determine which of the 14 predictors (6 POMS and 8 VAS) was significantly related to the criterion variable (RTSW), BMDP2R was used. A forward-stepping approach was employed in which the F to enter was set to 4.0. The entire data set (all sessions and all conditions) was used to determine the best predictors regardless of testing time (session) or intervention/ condition (Znap, Pnap, and Nonap). Next, a separate stepwise regression was performed on the data from each condition separately. In this analysis, the independent (predictor) variables, determined from the stepwise procedure on the entire data set were forced into the equation in the same order as was calculated by the stepwise regression performed on the entire data set.

Bivariate correlations

After the stepwise regression determined which of the 14 independent variables was most closely related to the RTSW score, the data were analyzed with BMDP8D to compute bivariate correlations between RTSW scores and scores from the subjective data. This was done within each of the three test conditions (Znap, Pnap, and Nonap) separately.

Discriminant analysis

A discriminant analysis on all the data considered as a single group was conducted to determine which subset of the 14 variables was most useful for classifying subjects into 3 levels of sleepiness based on subjective measures. The data were analyzed with BMDP7M and both a classification matrix and jackknifed classification were obtained.

Results

MANOVA

The MANOVA on all of the dependent measures across the 3 levels of conditions (Znap, Pnap, and Nonap) and the 16 levels of session (6 predeprivation and 10 deprivation) indicated a significant condition-by-session interaction (Wilk's $\lambda = 0.26$; $F(450,6727.68)=1.56$, $p<.0001$), a significant session main effect (Wilk's $\lambda = 0.06$; $F(225,2602.90)=3.58$, $p<.0001$), and a significant condition main effect (Wilk's $\lambda = 0.17$; $F(30,40)=1.93$, $p=.0261$). The interaction was

5

due primarily to significant univariate effects on POMS vigor and fatigue; VAS alertness, anxiety, energy, confidence, irritability, and sleepiness; and RTSW latency to stage 2 sleep (p<.05). Examples of these differences (for POMS vigor, VAS alertness and sleepiness, and RTSW sleep latency) are shown in figure 1. The session main effect (indicative of the impact of sleep deprivation across all conditions) was due to significant univariate effects on every dependent measure with the exception of VAS anxiety (p<.05). Generally, these effects indicated decrements in mood and alertness as a function of sleep deprivation. The condition main effect was attributable primarily to univariate results on VAS irritability, VAS sleepiness, and RTSW latency to sleep (p≤.05), although the presence of higher-order interactions confuses the interpretation of these. A complete description of these effects is beyond the scope of the present report, but the effects do indicate that the tests employed were sensitive to the test conditions and sleep deprivation. A more detailed discussion of the general univariate results is available in Caldwell & Caldwell (in press).
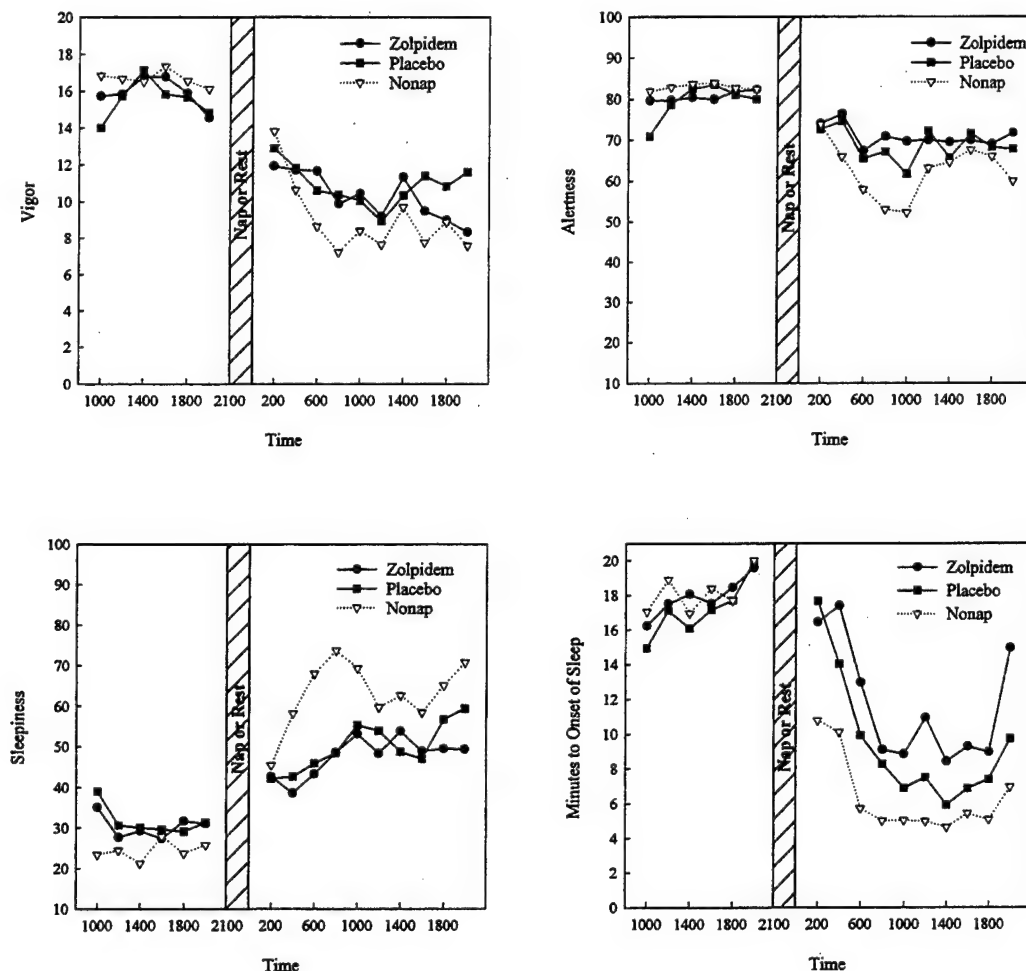

Figure 1. Interaction between condition and session (testing time) on
POMS vigor, VAS alertness, VAS sleepiness, and RTSW scores.

6

Multiple regression

The multiple regression on the entire data set (all sessions and all conditions, or 864 cases) indicated that 34 percent of the variance in RTSW scores (latency to stage 2 sleep) could be explained by the linear combination of POMS scores (tension, depression, anger, vigor, fatigue, and confusion) and VAS scores (alertness, anxiety, energy, confusion, irritability, jitteriness, sleepiness, and talkativeness) considered jointly. The multiple R of 0.59 was significant ($F(14,849)=31.769$, $p<.0001$), and the regression equation resulted in a standard error of estimate of 5.43. The predicted RTSWs are plotted against the actual RTSWs in figure 2.
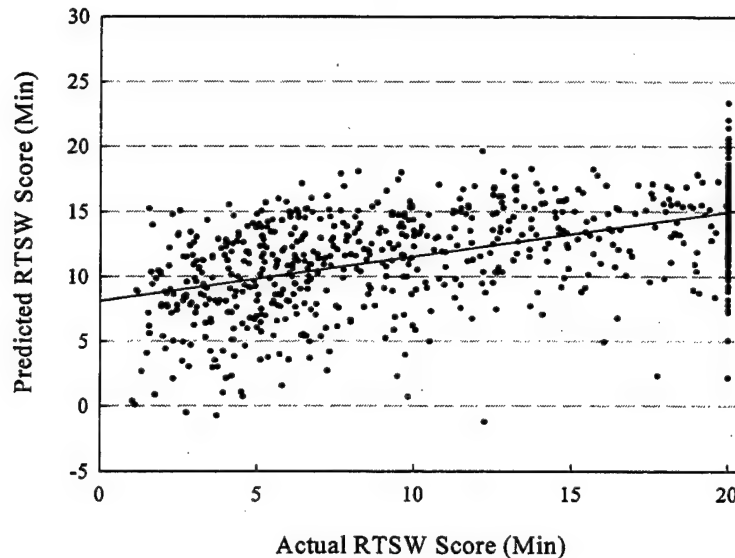

Figure 2. Predicted versus actual RTSW scores based
on the least squares multiple regression for
all conditions considered jointly.

The multiple regression for each of the conditions analyzed separately indicated the best prediction accuracy was in the Nonap condition, with the least accuracy in the Pnap condition. POMS and VAS scores after Nonap explained 53 percent of the variance in RTSW ($F(14,273)=22.076$, $p<.0001$), whereas subjective ratings after Pnap explained only 24 percent of the variance in RTSW ($F(14,273)=6.228$, $p<.0001$), and subjective ratings after Znap explained 32 percent of the variance in RTSW ($F(14,273)=9.180$, $p<.0001$). Predicted versus observed RTSWs for each separate condition are depicted in figure 3. A comparison of the slopes and intercepts of the regression lines indicated there was a significant difference among the three conditions ($F(30,819)=2.616$, $p=.00001$).

Stepwise multiple regression

The stepwise multiple regression on all 864 cases (or 864 sets of dependent variables) indicated that of the 14 POMS and VAS scores, only 7 contributed uniquely and significantly to
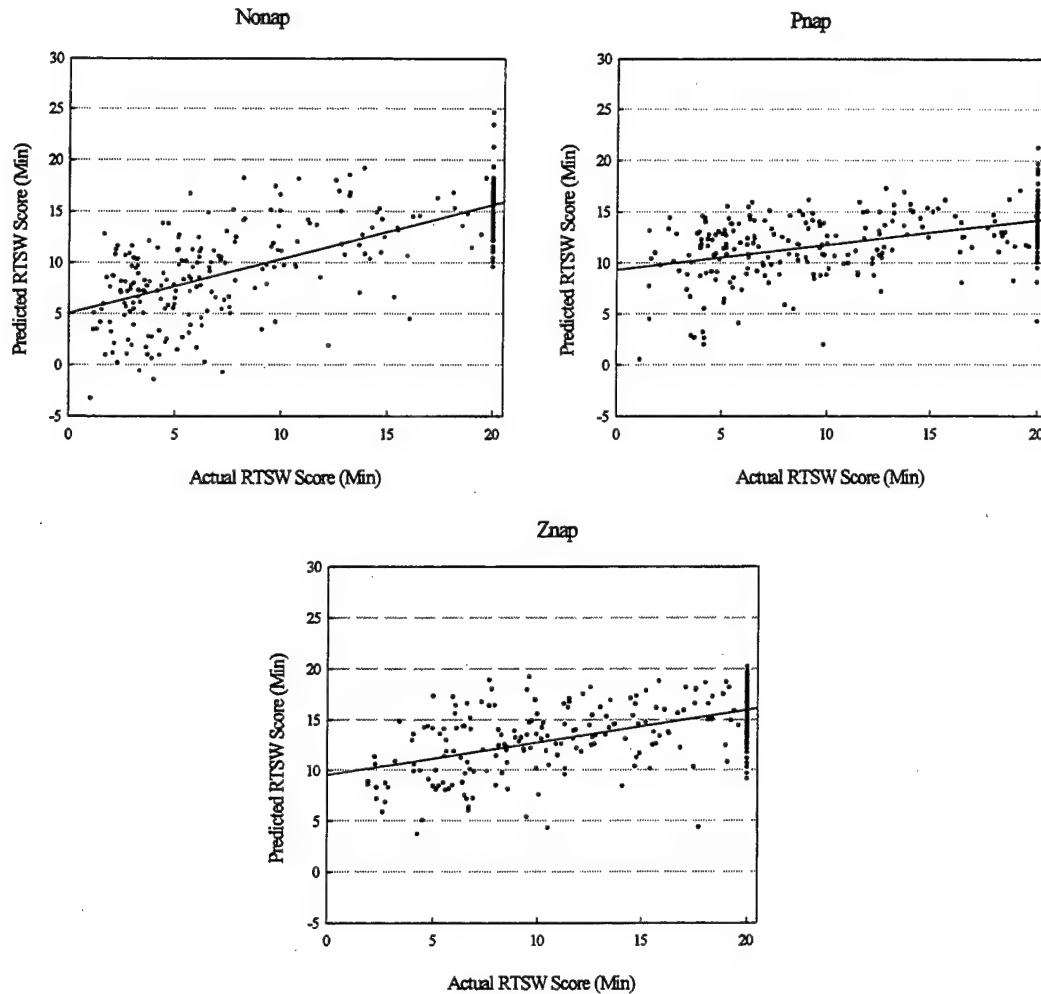
Figure 3. Predicted versus actual RTSW scores based on the least squares
multiple regression for the Nonap, Placebo nap and Zolpidem
nap conditions.

making an accurate prediction of sleepiness (as objectively measured by the RTSW). VAS
sleepiness alone explained 24 percent of the variance in RTSW scores ($F(1,862)=278.48$,
$p<.0001$), and POMS fatigue ($F(2,861)=173.20$, $p<.0001$), VAS alertness ($F(3,860)=125.01$,
$p<.0001$), VAS anxiety $F(4,859)=102.29$, $p=.0001$), VAS jitteriness ($F(5,858)=84.18$, $p<.0001$),
POMS depression ($F(6,857)=71.58$, $p=.0001$), and VAS talkativeness ($F(7,856)=62.24$, $p<.0001$)
together explained an additional 9 percent. Thus, the seven variables jointly explained a total of
33 percent of the variability in latency to stage 2 sleep in the RTSW. The stepwise regression
conducted for each of the three test conditions (using the same seven variables in the same order
for each) indicated that in the Nonap condition, the seven variables explained 52 percent of the
variance in RTSW. In the Pnap condition, the variables explained 23 percent of variance; and in
the Znap condition, the seven variables explained 29 percent of the variance.

8

Bivariate correlations

Within each of the three conditions, Pearson correlations were calculated between the RTSW and each of the variables extracted in the stepwise regression. The results are presented in table 1.

Table 1.
Correlation coefficients for the variables selected by the stepwise regression.

Condition	VAS Sleepy	POMS Fatigue	VAS Alert	VAS Anxiety	VAS Jittery	POMS Depression	VAS Talkative
All Groups	-0.49*	-0.47*	0.36*	-0.01	-0.09	-0.14	0.23
Znap	-0.39*	-0.35*	0.28*	-0.13	-0.11	-0.02	0.19
Pnap	-0.37*	-0.37*	0.26*	-0.03	-0.09	-0.22	0.17
Nonap	-0.67*	-0.61*	0.48*	0.12	-0.05	-0.28*	0.30*

* Significant $p<.01$ (the table value for r with 100 degrees of freedom for $p=.01$ is .254). Note that in this case, there are 862 degrees of freedom for all groups and 286 degrees of freedom for each of the individual conditions.

Table 1 shows there were substantial relationships between the first 3 variables (selected via the stepwise regression) and RTSW when all testing times were collapsed (16 sessions x 18 subjects, or 288 observations per condition). Note that bivariate correlations for all variables would not be expected to reach significance in this context since a straightforward product-moment correlation does not control for the impact of the other variables which were entered into the stepwise regression equation. However, it was of interest to determine whether within-cell correlations between the best self-report measures and the RTSW would attain significance despite the small sample size. When Pearson correlations were calculated within each of the individual cells (with 18 observations per cell), few significant effects were found. In fact, this procedure, which was carried out only on the VAS sleepiness and RTSW scores, indicated only 8 of the 48 individual tests were statistically significant at $p<.05$ ($r \geq 0.468$). In all eight instances, the relationship between VAS sleepiness and RTSW was negative, indicating that as sleepiness increased, latency to sleep decreased (in agreement with the overall correlations). The cells in which these relationships were found to be significant were as follows: Znap predeprivation

9

1000; Pnap predeprivation 1200; Znap deprivation 1000; Pnap deprivation 1400; and Nonap deprivation 1400, 1600, and 1800.

Discriminant analysis

The stepwise discriminant analysis selected the best combination of subjective variables to classify subjects into 3 groups of sleepiness (based on RTSW scores). The scores on VAS sleepiness, POMS fatigue, VAS anxiety, VAS alertness, POMS depression, and VAS jitteriness correctly classified subjects into groups of low, medium, and high alertness 58 percent of the time. A jackknifed classification using the same variables and groups correctly classified subjects into the groups 57 percent of the time. Classification in the high-alertness group was 71 percent accurate for subject placement. The medium-alertness had the lowest accuracy for correct classification into groups at 35 percent. Details of the discriminant analysis are shown in table 2.

Table 2.
Number of cases classified into low, medium, and high alertness groups.

CLASSIFICATION MATRIX*

Group	Percent correct	Number of cases classified into group		
		Low	Medium	High
Low	55.9	160	82	44
Medium	35.3	50	66	71
High	71.1	29	84	278
TOTAL	58.3	239	232	393

JACKKNIFED CLASSIFICATION MATRIX

Group	Percent correct	Number of cases classified into group		
		Low	Medium	High
Low	55.2	158	83	45
Medium	33.2	50	62	75
High	70.6	30	85	276
TOTAL	57.4	238	230	396

* In the classification matrix, there were actually 286 cases in the low group, 187 cases in the middle group, and 391 cases in the high group.

Discussion

Both subjective and objective measures of sleepiness/alertness were sensitive to the effects of sleep deprivation as shown by a significant overall multivariate session effect. This is

generally consistent with earlier reports in the literature (Hoddes et al., 1973; Mitler and Miller, 1996; Penetar et al., 1993; Kryger et al., 1991). However, the strength of the relationship between subjective reports and the RTSW was not particularly strong--a finding which also supports previous results from research with the MSLT (Johnson et al., 1991; Chervin et al., 1995; Seidel et al., 1987).

The relationships between RTSWs and self-reports were better under some conditions than others. When these relationships were examined for each of the three conditions separately (a 2-hour zolpidem nap, a 2-hour placebo nap, and a forced-rest condition), they were found to be best for the forced-rest condition and worst for the Pnap condition. This finding tends to support those of Johnson et al. (1991) who reported that treatments designed to increase or decrease sleepiness (hypnotics or caffeine) increased variability and weakened the correlations between subjective and objective measures of sleepiness/alertness. Within the Nonap condition, the wide differences in individual responses to total sleep deprivation may have increased the magnitude of the correlation coefficient by eliminating range restrictions in the values obtained from the various tests (Edwards, 1976).

Increasing the number of dimensions of subjective indicators apparently had little positive effect on the correlations between subjective and objective assessments. The use of all 14 of the VAS and POMS scores together explained only 34 percent of the variance in the RTSW. This level of prediction accuracy is only 10 percent better than the accuracy of predicting sleep latency based on a single VAS sleepiness scale. Because of low overall correlations among variables, it was not surprising to find that accurately classifying subjects into high-, medium-, and low-alertness groups (which were established based on RTSW scores) was difficult. Although overall classification accuracy was about 58 percent correct, cases which fell into the medium-alertness group were correctly classified only 33-35 percent of the time. This indicates that subjects who are not at the extremes of the sleep-deprivation continuum tend to be less self-aware than those who are clearly alert or very sleepy--a fact that probably accounts for much of the overall problem in obtaining high correlations between subjective and objective measures of sleepiness.

Thus, the overall results of this study showed that while instructional set may affect the dimension of sleepiness/alertness that is measured by objective sleep latency tests (Carskadon and Dement, 1982; Hartse, Roth, and Zorick, 1982), placing the focus on maintenance of wakefulness versus the initiation of sleep apparently had little impact on the strength of association between subjective and objective indicators of sleepiness/alertness. This is an interesting finding in view of the fact that Sugerman and Walsh (1989) reported the RTSW was more sensitive than the MSLT in situations similar to the present (i.e., the effects of napping versus no-napping on degree of sleepiness). Because of the potential increased sensitivity of the objective measure used here (the RTSW), one might have hypothesized that correlations between objective and subjective measures may have improved; however, this was not the case. In fact, the absolute values of correlations between RTSWs and the best self-reports ranged from only

11

0.49 to 0.36 (for VAS sleepiness, POMS fatigue, and VAS alertness). Correlations with the other 11 self-report scales were much lower.

Perhaps the results reported here could have been affected by the methods used to evaluate the relationships among the variables of interest. Specifically, the present study was limited by the fact that there are no regression or correlational procedures designed for repeated measures analyses, and it was not feasible to expose a different sample of volunteers to each level of sleep deprivation across the various interventions. However, because of earlier reports that time of testing exerts a significant impact on the relationship between subjective and objective assessments of sleepiness/alertness (Johnson et al., 1991), it was felt important to include observations from a variety of different times in order to accurately characterize the relationships of interest. Unfortunately, to accomplish this within the context of a completely between-subjects design would have required approximately 2400 volunteers if total independence of every observation was to be ensured (this is based on the assumption that 50 subjects per cell would have been ideal for testing the impact of sleep loss at 16 different times under each of 3 interventions for each of the measures examined). Instead, a repeated-measures approach was used in which the individual replicates from 18 subjects (exposed to all conditions) provided the data for analyses. This may have inflated the magnitude of any observed relationships since the measures were not independent, and although the statistical tests used here were made more stringent to compensate for this problem, an undetectable bias in the results may have occurred; however, this seems not to be a substantial cause for concern since the present results are similar to those of other investigators who used larger samples (although with the MSLT rather than the MWT or the RTSW). In the future, it may be possible to conduct a "more pure" analysis of these measures once the data from multiple similar studies can be collected and consolidated.

Until this is possible, the present results hopefully offer a reasonable estimate of the relationship between subjective measures of sleepiness/alertness and an objective measure of maintenance of wakefulness. Although rather weak relationships were observed, a single VAS rating of sleepiness can provide a basic indication of the desire for sleep that correlates significantly with the results of a maintenance of wakefulness test (the RTSW) in situations where effort and expense prohibit the use of the latter type of technique. Based on the present findings, the POMS fatigue scale also would be expected to contribute valuable information.

12

References

Alloway, C. E., Ogilvie, R. D., and Shapiro, C. M. 1997. The alpha attenuation test: Assessing excessive daytime sleepiness in narcolepsy-cataplexy. Sleep. 20: 258-266.

ASDA-Reports from the American Sleep Disorders Association. 1992. The clinical use of the multiple sleep latency test. Sleep. 15: 268-276.

Broughton, R. J. 1992. Qualitatively different states of sleepiness. Sleep, arousal and performance. R. J. Broughton and R. D. Ogilvie (eds). Boston: Birkhauser. 45-59.

Caldwell, J. A., and Caldwell, J. L. 1997 (in press). Comparison of the effects of zolpidem-induced prophylactic naps and forced rest periods in prolonged work schedules.

Carskadon, M. A., and Dement, W. C. 1982. The multiple sleep latency test: What does it measure? Sleep. 5: S67-S72.

Carskadon, M. A., and Roth, T. 1994. Normal sleep and its variations. Principles and practice of sleep medicine. M. H. Kryger, T. Roth, and W. C. Dement (eds). Philadelphia: W. B. Saunders Company.

Chervin, R. D., Kraemer, H. C., and Guilleminault, C. 1995. Correlates of sleep latency on the multiple sleep latency test in a clinical population. Electroencephalography and clinical Neurophysiology. 95: 147-153.

Clodore, M., Foret, B. J., and Bouard, G. 1990. The multiple sleep latency test: Individual variability and time of day effect in normal young adults. Sleep. 13: 385-394.

Edwards A. L. 1976. An introduction to linear regression and correlation. San Francisco: W. H. Freeman and Company.

Elsmore, T. F., Hegge, F. W., Naitoh, P., Kelly, T., Schlangen, K., and Gomez, S. 1995. A comparison of the effects of sleep deprivation on synthetic work performance and a conventional performance assessment battery. San Diego: Naval Health Research Center. NHRC Report No. 95-6.

Hartse, K. M., Roth, T., and Zorick, F. J. 1982. Daytime sleepiness and daytime wakefulness: The effect of instruction. Sleep. 5: S107-S118.

Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., and Dement, W. C. 1973. Quantification of sleepiness: A new approach. Psychophysiology. 10: 431-436.

Johns, M. W. 1991. A new method for measuring daytime sleepiness: The Epworth sleepiness scale. Sleep. 14: 540-545.

Johnson, L. C., Freeman, C. R., Spinweber, C. L., and Gomez, S. A. 1991. Subjective and objective measures of sleepiness: Effect of benzodiazepine and caffeine on their relationship. Psychophysiology. 28: 65-71.

Kryger, M. H., Steljes, D., Pouliot, Z., Neufeld, H., and Odynski, T. 1991. Subjective versus objective evaluation of hypnotic efficacy: Experience with zolpidem. Sleep. 14: 399-407.

McNair, D. M., Lorr, M., and Droppleman, L. F. 1981. Manual for the profile of mood states. San Diego: Educational and Industrial Testing Service.

Mitler M. M., Gujavarty, K. S., and Browman, C. P. 1982. Maintenance of wakefulness test: A polysomnographic technique for evaluating treatment in patients with excessive somnolence. Electroencephalography and clinical Neurophysiology. 53: 648-661.

Mitler, M. M., and Miller, J. C. 1996. Methods of testing for sleeplessness. Behavioral Medicine. 21: 171-183.

Penetar, D., McCann, U., Thorne, D., Kaminori, G., Sing, H., Thomas, M., and Belenky, G. 1993. Caffeine reversal of sleep deprivation effects on alertness and mood. Pharmacology. 112: 359-365.

Roth, T., Hartse, K. M., Zorick, F., and Conway, W. 1980. Multiple naps and the evaluation of daytime sleepiness in patients with upper airway sleep apnea. Sleep. 3: 425-439.

Sangal, R. B., Thomas, L., and Mitler, M. M. 1992. Maintenance of wakefulness test and multiple sleep latency test: Measurement of different abilities in patients with sleep disorders. Chest. 101: 898-902.

Seidel, W. F., Cohen, S., Bliwise, N. G., and Dement, W. 1987. Cetirizine effects on objective measures of daytime sleepiness and performance. Annals of Allergy. 59: 58-62.

Sugerman, J. L., and Walsh, J. K. 1989. Physiological sleep tendency and ability to maintain alertness at night. Sleep. 12: 106-112.

Thorne, D. R., Genser, S. G., Sing, H. C., and Hegge, F. W. 1985. The Walter Reed performance assessment battery. Neurobehavioral toxicology and teratology. 7: 415-418.

Walsh, J. K., Muehlbach, M. J., Humm, T. M., Dickins, Q. S., Sugerman, J. L., and Schweitzer, P. K. 1990. Effects of caffeine on physiological sleep tendency and ability to sustain wakefulness at night. Psychopharmacology. 101: 271-273.

14

Wilkerson, R. T. 1961. Interaction of lack of sleep with knowledge of results, repeated testing, and individual differences. Journal of Experimental Psychology. 62: 263-271.

Wilkerson, R. T. 1964. Effects of up to 60 hours' sleep deprivation on different types of work. Ergonomics. 17: 175-186.

Wilkerson, R. T. 1965. Sleep deprivation. The physiology of human survival. O. G. Edholm and A. L. Bacharach (eds). London: Academic Press. 399-430.

Wilkerson, R. T. 1969. Some factors influencing the effect of environmental stressors upon performance. Psychological Bulletin. 72: 260-272.